

Prospective nucleotide sequence analysis of methicillin resistant *Staphylococcus aureus* isolates from Sokoto state

Shuaibu Suleiman Adeiza, Josiah Ademola Onaolapo, Busayo Olalekan Olayinka

Department of Pharmaceutics and Pharmaceutical Microbiology, Faculty of Pharmaceutical Sciences, Ahmadu Bello University, Zaria, Kaduna, Nigeria

Summary

The epidemiology of MRSA has evolved considerably and a better understanding of the reservoir mix and phylogenetic resolutions of MRSA may inform strategies to prevent MRSA transmission and infection. Even-though in Nigeria there are studies that have reported on the molecular characterization of MRSA, sadly, not much effort has been put into analysis and submission of sequenced data to Gene repositories. To the best of our knowledge, this study is the first of its kind in Sokoto to have successfully deposited nucleotide sequences of MRSA strains into Genbank database. The sequence reads of molecularly characterized strains should be deposited into Gene repositories by researchers to further populate the databank and improve the efficiency of local phylogenetic analysis to foster better bioinformatics as paucity of local data exists in these databases.

Correspondence: Shuaibu Suleiman Adeiza, Department of Pharmaceutics and Pharmaceutical Microbiology, Faculty of Pharmaceutical Sciences, Ahmadu Bello University, Zaria, Kaduna, Nigeria.
E-mail: suleykestler2@gmail.com

Key words: MRSA, Phylogeny, Cladogram, Sokoto.

Contributions: The authors contributed equally.

Conflict of interest: The authors declare no potential conflict of interest.

Funding: None.

Availability of data and materials: The sequences analysed here are accessible in GenBank®, (<https://www.ncbi.nlm.nih.gov/genbank/>).

Ethics approval and consent to participate: Ref. No. SMH/1580/V. IV.

Received for publication: 17 August 2019.

Revision received: 5 January 2020.

Accepted for publication: 20 January 2020.

©Copyright: the Author(s), 2020
Licensee PAGEPress, Italy
Microbiologia Medica 2020; 35:8496
[doi:10.4081/mm.2020.8496](https://doi.org/10.4081/mm.2020.8496)

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License (by-nc 4.0) which permits any non-commercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Introduction

Methicillin-resistant *Staphylococcus aureus* (MRSA) is a pathogen that is extremely impacting national and global health and has developed into a severe health burden (5). It is heralded by symptomless nasal carriage which may cause health difficulties in immunocompromised individuals (8). It has a remarkable ability to cause infections; from benign skin diseases to deep bone marrow infections (13).

The epidemiology of MRSA has evolved considerably. In the past, MRSA strains could be trailed back to a hospital or clinical source, but over the years there appears to be a decrease in hospital-acquired MRSA and an upsurge in the occurrence in community acquired strains (2). This may be due to improve hospital policies on MSA surveillance. However, the community-acquired strains have demonstrated more virulence than the hospital-acquired strains (1).

Evolutionarily, MRSA arose as an effect of selective pressure engendered by over exposure to antibiotics (3). It has a remarkable ability to respond to environmental pressures. This facultative anaerobe can withstand strict environmental settings such as high temperatures, salt concentrations and dry environmental conditions (15).

Methods based on molecular techniques are gradually becoming less expensive (adjusted for inflation rate) and more prevalent in Nigeria with molecular companies now in a quest for cost-effective ways of collaborating with molecular laboratories and researchers.

Even-though in Nigeria there are studies that have described the molecular characterization of MRSA. Sadly, not much effort has been put into submission of sequenced data into Gene repositories. Sequence deposition into gene repositories fosters better bioinformatics like; phylogenetics and specific primer design. A better understanding of the reservoir mix and diversity of MRSA may inform strategies to prevent MRSA transmission and infection in Sokoto. Bioinformation is of particular importance in the management MRSA infection in patients under intensive care because routine biochemical tests are slow and often fail to characterize the pathogenic strain due to phenotypic divergence.

The goal of the study is to recognize the phylogenetic affiliation among MRSA strains recovered from Sokoto State-owned hospitals with a view to ascertain the genetic diversity of the hospital strains based on *16S rRNA* and *mecA* gene sequencing.

Materials and methods

Study center

The present research work was carried out on strains from Sokoto State, located at the extreme north-west of Nigeria, between

longitudes 4°8' and 6°54' and latitude 12°N and 13°58'N. The state has a population of nearly 5.4 million people (14), covers a terrestrial area of 32,000km² and shares a border with the Republic of Niger. The major ethnic groups in the state are the Hausa and Fulani groups. Over 80 percent of the people in the state practice agriculture as their major source of income while local crafts like black smiting, weaving, dyeing, carving and leather works form the remaining 20% (16).

Bacterial strains

In this study, 16 *mecA* positive MRSA clinical strains obtained from the nostrils of participants from Sokoto state-owned hospitals were selected for sequencing. All the enrolled strains were previously characterised using standard microbiological procedures (Gram stain, Catalase and Coagulase test), Microgen™ staph identification Kit, DNA extraction and PCR amplification. Additional characterization was carried out on these strains by DNA sequencing and phylogenetic analysis.

Staphylococcal DNA extraction and PCR amplification

Staphylococcal genomic DNA was harvested using Qiagen (Hilden, Germany) DNA extraction kit in adherence to the manufacturer's protocols. For each strain, DNA fragment (about 956 bp) of the *16SrRNA* and *mecA* (533 bp) genes were amplified according to methods previously described by Hassan (7).

Sequencing

The amplicons were purified using Monarch® bio-labs DNA clean-up kits according to the manufacturers' recommendations. The purified PCR products was packaged and sent for Sanger sequencing at Inqaba Biotec South Africa.

Bioinformatics and DNA Sequence Analysis

The chromatograph sequence files were analyzed with the software package Mega™ 7 (version 7.0.26). After providing the input sequence (ABI format) the software was used to mask out low quality sequence reads. The nucleotide sequences were converted into FASTA file format for further analysis.

Gene annotation

The FASTA formatted nucleotide sequences were marked for their Open reading frames (ORF), Coding sequences (CDS), Untranslated regions (UTR), START codons (ATG), STOP codons (TAA), Promoter region (-35, -10 and shine-Dergano sequences) and the terminator earpin using CLC genomic work bench software® (Version 3.6.5).

Nucleotide blast

The annotated sequences were entered into the Basic Local Alignment Search Tool (BLASTN) of National Center for Biotechnology Information (NCBI) database to assess the quality of the sequences. Sequences alignments from the 5' to 3' ends was also determined on the platform. The E-value, topology, percentage G and C component, pairwise identities and sequence length were also determined.

Phylogenetic analysis

All sequences were imported into Geneious® prime software (Bio-matters Ltd) program to further assess the query nucleotides in a customized environment and thus calculate the phylogenetic tree using the genetic distance model and a tree build method of neighbour joining neighbour. Pairwise comparison was used to rank the strains based on their phylogenetic affiliation; a comparison of

row items (% identity) with column item (genetic distance). The smaller numbers in the comparison reveals sequences that are closer together or has a more recent common ancestor and vice-versa. The strains were assigned identification numbers 1 to 12 for convenience. The diagonal in the chart is an intercept of "like for like" where no comparisons were carried out. A combined Metatree (<https://itol.embl.de>) was further calculated in order to visualize the evolutionary relationships of the strains within a broader framework.

Sequence submission to GenBank

The source modifiers, the sequences of the primer region masked out, the sequencing methods (in this case Sanger) and other features of the nucleotide sequences were distinguished before they were submitted to GenBank for processing. The nucleotide sequences were submitted as follows; 16S ribosomal DNA partial CDS were submitted directly into GenBank repository and *mecA* gene partial CDS were submitted via Bankit. After full processing by gene bank Staffs and accepted for publication in their repository, accession numbers were issued as identifiers for each submitted sequence and a release date is pronounced depending on the queue. A FLATFILE was also supplied with a detailed preview of the GenBank format before publication.

Results

The submitted sequence data are accessible at GenBank under accession MK643256- MK643267 (*16SrRNA*) and MK659546- MK659557 (*mecA*). Figures 1-3 shows phylogenetic trees constructed after sequences were aligned with each other and with other strains from GenBank.

The cladogram (Figure 2) shows that 8 out of 12 (66.7%) of the MRSA strains are monophyletic and clustered together in a Single

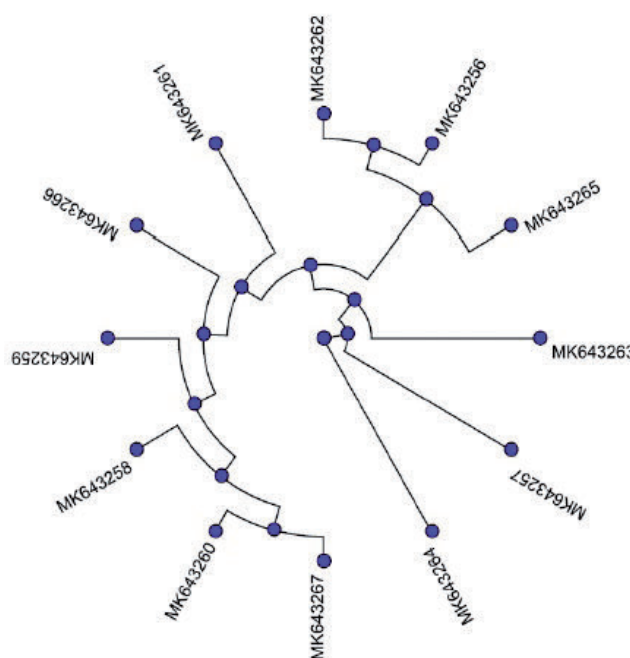


Figure 1. Circular phylogenetic layout of strains sequenced for *16srRNA*.

clade. In the Figure, 25.0% (3 out of 12) of the sequences were polyphyletic and 8.3% of the strains are paraphyletic and appeared to have arisen from a distant ancestor relative to others.

The genetic relationships among the sequences of 12 MRSA strains with Sequences (in terms of their *mecA* gene) from GenBank were calculated in a cladogram (Figure 3). Based on the phylogeny, the strains were grouped in to five clades. Two different clades contained 3 strains (25% each) and other two clades clustered 2 MRSA (16.7%) strains each. The cladogram spreads radially with the first clade (clockwise direction) occurring furthest away in the evolutionary tree.

Table 1 shows that the strain MK643264 was the most distant, with low similarity of 4.25% to 72% and an evolutionary distance of 0.35 to ∞ , from all other strains. The next distant strain MK643261, exhibited 70.58%- 92.43% similarity and approximately 0.08-0.37 evolutionary distance from other strains. Sequences like MK643256, MK643258, MK643260, MK643262, MK643263, MK643265, MK643266 and MK643267 seemed to be grouped with higher similarities (red spectrum). The sequence groups shared 60.34-90.31% similarity with an evolutionary distance of 0.08-1.94.

The native strains of the major evolutionary cluster in Figure 2 appear to populate the red spectrum in the lower left side of Table

2 with high percentage similarity of 97.38% to 100% with little or no evolutionary distance from each other (0.00 to 0.3).

Figure 4 shows the difference amongst nucleotides sequences that were aligned. This difference was as low as one nucleotide per 100 residues and high as 100 nucleotides per wrap.

Discussion

Policies to control MRSA infections in hospitals and community environments require baseline knowledge on the genetic diversity of circulating strains. This study provides information on the analysis of local MRSA strains of unknown diversity collected from a hospital environment with the aim of uncovering otherwise undetected differences that may aid infection control strategies.

All 24 sequences (*16srRNA*(n=12) and *mecA* (n=12)) in this study were queried through a blast search and alignment for similarity among the local strains. Query sequences were evaluated and compared with the GenBank database sequences based on the following criteria: query coverage, e-value, and percent identity. Query coverage represents the percentage of query sequence length

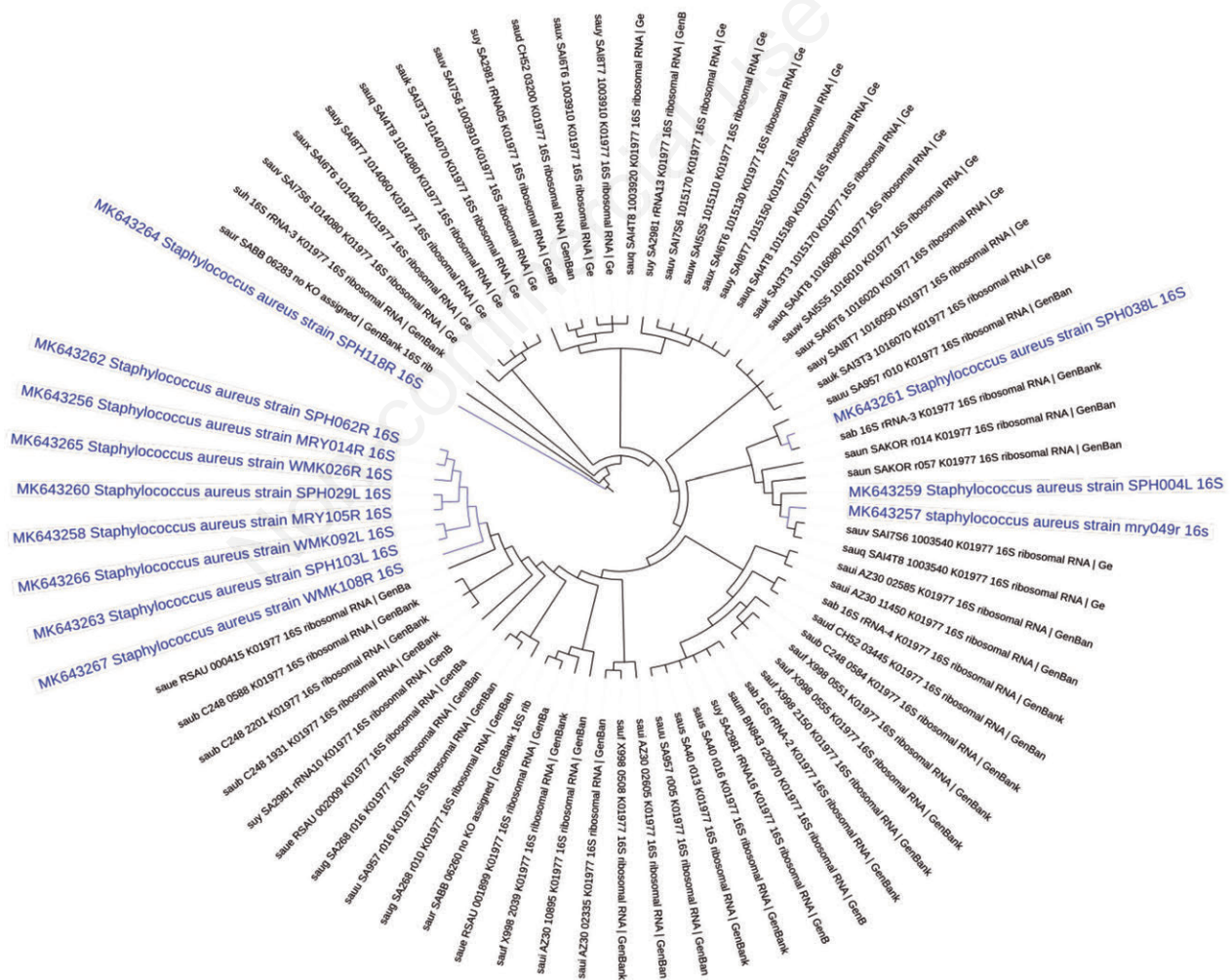


Figure 2. Cladogram map based on genetic relatedness of 16srRNA gene of 12 MRSA strains with 60 query strains from GenBank and Genome.Net database. The positions of the twelve MRSA strains are indicated in blue within the circle. Circular representation of the phylogenetic tree obtained using Multiple Sequence Comparison by Log-Expectation (MUSCLE) algorithm.

that is included in the alignment. In this study the query coverages ranged from 40-100%, mostly due to the fact that most of the sequences obtained were partial and their coverage, when aligned with whole genome sequences was low.

The e-value is a parameter that represents the number of chance alignment and directly dependent on the size of the database, such that a smaller database has higher likelihood of containing the sequence of interest. Search queries were restricted to the database of the genes of interest. This filtering procedure leads to an e-value of zero (0) for all sequence alignment in this study.

The percent identity refers to the part of residues in the compared nucleotide sequences that match at the same nucleotide position. The percentage identical sites in this study varied from 17.8-100% for *16sRNA* and 13.65%- 100% for *mecA*. These

findings agree with a previous work (11) that reported that organisms that have fewer than 97.0% homology of *rRNA* will not associate to more than 60% of genomic relatedness. The alignment was carried out to understand the sequence variations in order to estimate the relatedness for the strains. In this study we noticed nucleotide differences at various points in the alignment which may lead to a change in the amino acid sequences or phenotypic traits.

The evolutionary histories of 12 isolates based on *16SrRNA* and *mecA* sequences were inferred using the multiple sequence comparison by log expectation (MUSCLE) algorithm. The constructed phylogenetic tree clearly resolved the relationships among the sequences. The tree was calculated to scale, with branch lengths having identical measurements as those of the evolutionary distances used to generate the phylogenetic tree. In the case of the

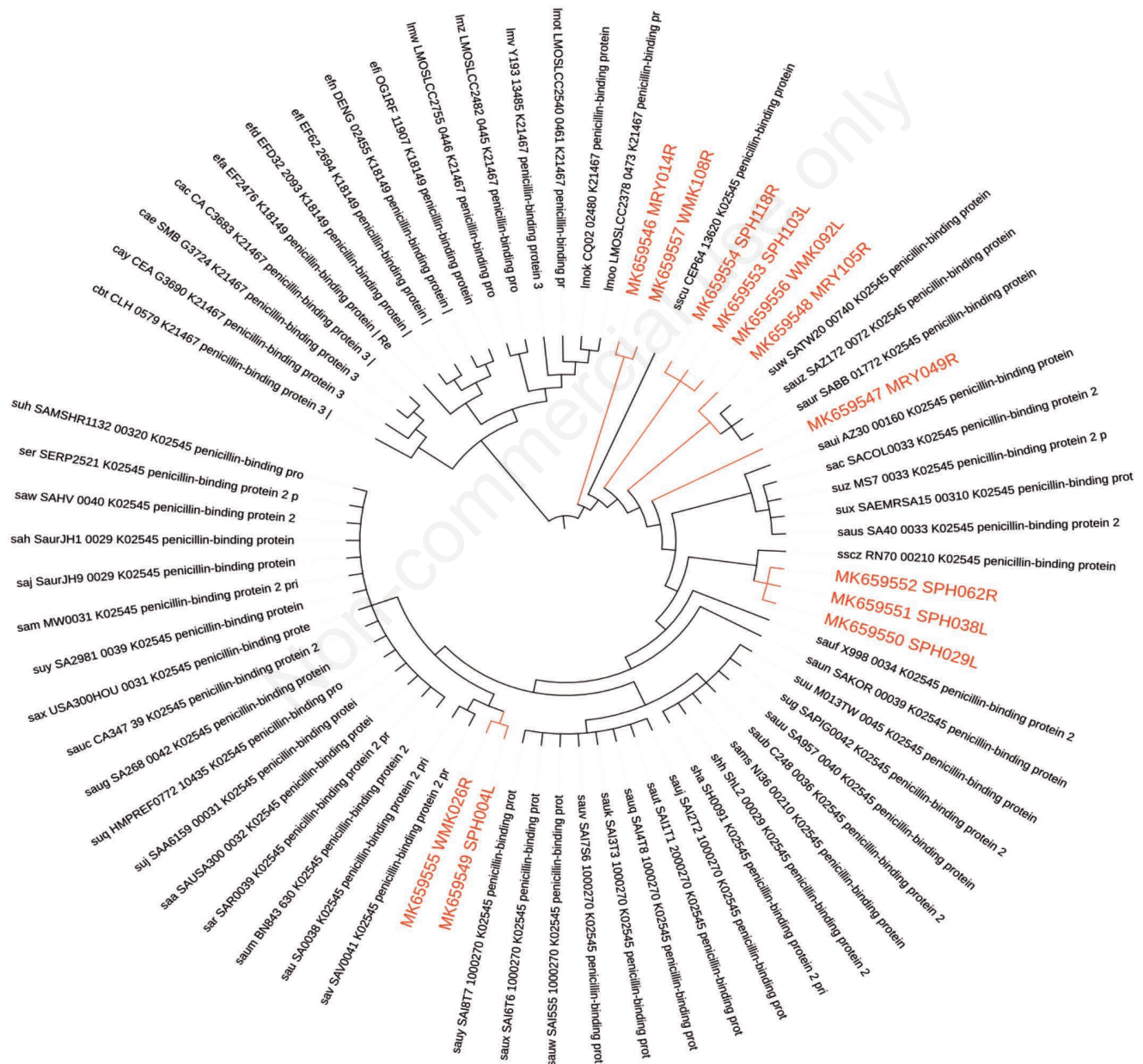


Figure 3. Cladogram based on genetic relatedness *mecA* gene of 12 MRSA strains with 60 query strains from GenBank and Genome Net database. The positions of the twelve MRSA strains are indicated in red within the circle. Circular representation of the phylogenetic tree obtained using Multiple Sequence Comparison by Log-Expectation (MUSCLE) algorithm.

16SrRNA the sequences were clustered into 4 clades and *mecA* into five clades in accordance with their genetic similarities.

The tree depicts the overall phylogeny of strain across the species. The root of the tree indicates the ancestral state for the entire phylogeny. The tree data structure expands outwards and the length of each path increases with the radius (spirals outwards from the root of the tree). The number of nodes multiplies exponentially from the first node while the circumference increases linearly resulting in a tree with the external circumference packed with nodes (taxa or strains).

The data presented in the meta-tree shows that most (66.7%) of the MRSA strains were monophyletic based on their *16SrRNA* genes and supports the claim that the gene is highly conserved among bacteria strains (10). Conversely, the diversity of the outstanding 33.3% suggests that conserved regions of the gene are probably not as conserved as anticipated, which is in agreement with a previous

finding (9) where the conserved regions of the *16SrRNA* gene revealed important variation between phylogenetic groups. However, diversities observed in this study may be due to differences in the DNA extraction and ribosequencing methods employed by the gene depositors. The mounting cases of horizontal transfer of *16SrRNA* genes (Genetic promiscuity) among different bacterial strains reported by Miyazaki (12) also explicate our findings. Furthermore, mutation in the hypervariable region of 16S ribosome due to prolonged exposure to the antibiotics in the hospital environment, may also have contributed to the genetic diversity realized among the phylogenetic groups.

S. aureus strains morphs upon acquisition (MRSA) and excision (MSSA) of staphylococcal cassette chromosome *mec* or *SCCmec* (4). These variations have occurred multiple times in the evolutionary clock of *S. aureus* and may be responsible for the *mecA* gene-based diversities observed in this study.

Table 1. Pairwise comparison of Percentage identity (lower left) and genetic distance (Upper right) of twelve 16SrRNA sequences of MRSA.

	1	2	3	4	5	6	7	8	9	10	11	12
MK643260	1	0.29	0.10	0.16	0.19	0.11	0.22	0.35	0.23	1.02	1.92	∞
MK643267	2	76.23	0.36	0.22	0.49	0.37	0.32	0.72	0.54	0.48	0.95	1.88
MK643266	3	90.31	71.49	0.22	0.17	0.08	0.22	0.34	0.20	1.21	2.60	∞
MK643256	4	85.30	80.69	80.60	0.35	0.24	0.18	0.55	0.39	0.91	1.73	∞
MK643265	5	83.57	64.15	84.78	71.84	0.17	0.34	0.20	0.18	1.55	∞	∞
MK643261	6	90.20	70.58	92.43	79.78	85.06	0.23	0.33	0.19	1.26	2.83	∞
MK643263	7	81.18	74.15	81.02	83.97	72.38	80.13	0.56	0.39	0.96	1.85	∞
MK643258	8	71.73	53.83	72.48	60.90	82.30	73.40	60.34	0.20	∞	∞	∞
MK643262	9	80.00	61.34	82.21	69.52	84.32	83.11	69.41	82.48	1.94	∞	∞
MK643257	10	44.15	64.39	39.93	47.16	34.54	39.07	45.82	23.87	30.68	0.53	1.16
MK643259	11	30.76	46.22	27.36	32.43	23.58	26.72	31.40	14.70	20.04	62.00	0.35
MK643264	12	17.80	31.09	14.53	18.49	11.55	14.14	17.56	4.25	8.31	40.95	72.00

Key: Comparison gradient runs from Purple (min) to Red (max)



Table 2. Pairwise comparison of Percentage identity (lower left) and genetic distance (Upper right) of twelve mecA genes sequences of MRSA.

	1	2	3	4	5	6	7	8	9	10	11	12
MK659547 mecA	1	0.01	0.01	0.01	0.00	0.02	0.02	∞	∞	∞	∞	∞
MK659553 mecA	2	99.44	0.00	0.00	0.01	0.03	0.03	∞	∞	∞	∞	∞
MK659554 mecA	3	99.44	100.00	0.00	0.01	0.03	0.03	∞	∞	∞	∞	∞
MK659556 mecA	4	99.44	100.00	100.00	0.01	0.03	0.03	∞	∞	∞	∞	∞
MK659548 mecA	5	99.81	99.25	99.25	99.25	0.02	0.02	∞	∞	∞	∞	∞
MK659546 mecA	6	97.76	97.38	97.38	97.38	97.57	0.00	∞	∞	∞	∞	∞
MK659557 mecA	7	97.76	97.38	97.38	97.38	97.57	100.00	∞	∞	∞	∞	∞
MK659550 mecA	8	22.40	22.14	22.14	22.14	22.54	21.86	21.86	0.00	0.00	∞	∞
MK659551 mecA	9	22.40	22.14	22.14	22.14	22.54	21.86	21.86	100.00	0.00	∞	∞
MK659552 mecA	10	22.40	22.14	22.14	22.14	22.54	21.86	21.86	100.00	100.00	∞	∞
MK659549 mecA	11	20.13	19.87	19.87	19.87	20.13	20.40	20.40	13.65	13.65	13.65	0.00
MK659555 mecA	12	20.13	19.87	19.87	19.87	20.13	20.40	20.40	13.65	13.65	13.65	100.00

Key: Comparison gradient runs from Purple (min) to Red (max)

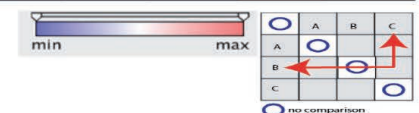




Figure 4. Multiple sequence alignment of twelve (12) nucleotide sequences of the 16S rRNA gene of MRSA isolates from Sokoto. The coloured bases indicate regions of nucleotides diversity among isolates. The identical residues are represented with dots. The sequence layout is set to wrap at every 100 residues.

The distance between the strains is a measure of the genetic distances between the strains. It is also a measure of times since the isolates diverge in evolution (17). Consequently, the strains with GenBank accession Identities MK643264 (*16SrRNA*), MK659555 and MK659549 are more distantly related to all other strains in the tree. The high nucleotides similarity between *16S rRNA* genes of strains that originated from racing horse (SA17S6, Korea) and human (Blood and Pus form RSAU000415, Japan, SA957, Australia and SABB06263, Brazil) gives credence to the probability of human-to-animal host jump events (6).

It can be argued from the tree that a number of diversities have taken place among the Sokoto-strains. The tree shows that the capacity to mutate has evolved on several occasions during the evolutionary clock of the strains. This may be due to mutations owing to selective pressures (Antimicrobial resistance) or human-to-animal host jump events. These jumps are not startling because *S. aureus* has a clonal population structure, it grows and diversifies by mutation more often than by recombination (horizontal acquisition of short stretches of DNA and between different strains which then recombine and replace existing stretches of DNA).

Conclusions

To the best of our knowledge, this study is the first of its kind in Sokoto to have successfully deposited nucleotide sequences of studied MRSA strains into Genbank database. It has also revealed the phylogenetic diversity of the strains. Based on sequence analysis it was found that a great deal of diversity has taken place among Sokoto-strains identified in this study. The sequence reads of molecularly studied strains should be deposited into GenBank database by researchers to further populate the database and improve the efficiency of local phylogenetic analysis to foster better bioinformatics as paucity of local data exists in these databases.

References

- Akinjogunla OJ, Ajayi OA, Ekeh NO. Virulence factors and Antibiotic Resistant *Staphylococcus* spp from the Anterior Nares of Apparently Healthy Undergraduate Students in Uyo. *Am J Res Commun* 2014;2:158-80.
- Calfee DP. Trends in Community Versus Health Care-Acquired Methicillin-Resistant *Staphylococcus aureus* Infections. *Curr Infect Dis Rep* 2017;19:48.
- Dongxin L, Qianting O, Jialing L, et al. A meta-analysis of the rates of *Staphylococcus aureus* and methicillin-resistant *S. aureus* contamination on the surfaces of environmental objects that health care workers frequently touch. *Infect Genet Evolut* 2014;563-74.
- Gajdacs M. The Continuing Threat of Methicillin-Resistant *Staphylococcus aureus*. *Antibiotics* 2019;8:52.
- Gordon RJ, Lowy FD. Pathogenesis of Methicillin-Resistant *Staphylococcus aureus* Infection. *Clin Infect Dis* 2008;46: S350-9.
- Harrison EM, Paterson GK, Holden MT, et al. Whole genome sequencing identifies zoonotic transmission of MRSA isolates with the novel *mecA* homologue *mecC*. *EMBO Mol Med* 2013;5:509-15.
- Hassan MM, Ismail IA. Isolation and molecular characterization of some pathogenic mobile phone bacteria. *Int J Biochem Biotechnol* 2014;3:516-22.
- Isibor JO. Asymptomatic Carriage of Plasmid-Mediated Multidrug-Resistant *Staphylococcus aureus* in Nasal Tracts of Persons in a Semi-Urban Area of Nigeria. *Am J Life Sci* 2014;2:200.
- Kabbashi AS, Hassan AM, Garbi MI, et al. A Novel Mutation in the Conserved Region of 16S rRNA Genes of *Escherichia coli* Clinical Isolates. *J Appl Sci* 2019;19:406-12.
- Kai S, Matsuo Y, Nakagawa S, et al. Rapid bacterial identification by direct PCR amplification of 16S rRNA genes using the MinION™ nanopore sequencer. *FEBS Open Bio* 2019;9:548–57.
- Lan Y, Rosen G, Hershberg R. Marker genes that are less conserved in their sequences are useful for predicting genome-wide similarity levels between closely related prokaryotic strains. *Microbiome* 2016;4:18.
- Miyazaki K, Tomariguchi N. Occurrence of randomly recombined functional 16S rRNA genes in *Thermus thermophilus* suggests genetic interoperability and promiscuity of bacterial 16S rRNAs. *Sci Rep* 2019;9:11233
- Morris DO, Loeffler A, Davis MF, et al. Recommendations for approaches to methicillin-resistant staphylococcal infections of small animals: diagnosis, therapeutic considerations and preventative measures. *Vet Dermatol* 2017;28:304.
- Nigeria Population and Development: Sokoto State Fact Sheet [Internet]. Sokoto State Government: Health policy plus; 2017 Sep p. 2. Report No.: No. AID-OAA-A-15-00051. Available from: www.healthpolicyplus.com
- Okwu M, Bamgbala S, Aborisade W. Prevalence of nasal carriage of community-associated Methicillin-resistant *Staphylococcus aureus* (CA-MRSA) among healthy primary school children in Okada, Nigeria. *Prevalence* 2012;2.
- Sokoto (State, Nigeria) - Population Statistics, Charts, Map and Location [Internet]. 2019 [cited 2019 Nov 1]. Available from: <https://citypopulation.de/php/nigeria-admin.php?adminid=NGA034>
- Truong DT, Tett A, Pasolli E, et al. Microbial strain-level population structure and genetic diversity from metagenomes. *Genome Res* 2017;27:626-38.